Estimation of correlation between soil organic matter and soil components using neural network

*Masaki Kiyohiro[1], Yasushi Mori[1], Yutaro Yamamoto[1]

1. Okayama University Graduate school of Environmental and Life Science

Soil is the terrestrial largest carbon storage, and, it is key to predict the carbon storage and loss. However, there are many estimation models, and we still cannot understand natural processes. Therefore, instead of predicting the organic matter from the conventional function system, we thought about letting the neural network learn anew and infer the feature value closely related to the soil organic matter.

Deep learning is a type of machine learning and it is also called a neural network. A neural network is composed of a unit called a node or a layer. The node receives a weighted input, outputs the weighted input to a next layer with a specific weight. After passing through all layers, output data can be obtained. The feature of deep learning is that the weight between each node can be selected other than 0 and 1, and other judgment criteria can be selected. Moreover, by preparing a plurality of hidden layers, it is possible to predict more complicated phenomena.

In this study, we conducted an experiment to predict the amount of organic matter in soil, and to estimate the correlation between organic matter and soil constituents using a neural network.

The surface soil data of HWSD (Harmonized World Soil Database) was used as data for learning. The target areas were Europe, Asia and Australia. The input items (features) are 17 items such as particle size distribution and cation exchange capacity.

In addition, feature selection was performed as data preprocessing. Feature selection is a method of automatically selecting important features to create a highly accurate model. In this experiment, the recursive feature elimination method (RFE) was used. Among the estimation models using the features, the one with the highest accuracy was obtained and its performance was evaluated. The model was used to estimate the effect of the selected feature on the amount of organic matter by changing each feature.

Among the features selected by the high-precision model created from the data of the target area, those that matched in three areas were silt content, dry density, cation exchange capacity of soil, and base saturation.

In addition, a model was created again using the features and parameters selected by the model with the highest accuracy, and the effect on organic matter by increasing or decreasing the value of a specific feature by 10% from the average value was determined. As a result, the impact was confirmed in Europe and Australia.

The experimental results suggest that the base model has a positive effect on the European model and a negative effect on the Australian model. In the Asian model, soil organic matter did not change significantly even when each feature was changed.

Prediction of soil organic matter and correlation estimation were performed using neural networks. Four common features were selected in the high-precision model generated from the data of three regions.

However, it was suggested that the effect of selected features on the amount of organic matter differed from region to region.

Keywords: Neural network , Soil organinc carbon, Correlation estimation

